Nucleic Acids Research, 1994, Vol. 22, No. 19 3997–4001

Internal fluorescence labeling with fluorescent deoxynucleotides in two-label peak-height encoded DNA sequencing by capillary electrophoresis

Heather R.Starke, Ju Ying Yan, Jian Zhong Zhang, Klaus Mühlegger[1], Klaus Effgen[1] and Norman J.Dovichi*

Department of Chemistry, University of Alberta, Edmonton, Alberta T6G 2G2, Canada and [1]Boehringer Mannheim GmbH, R&D Biochemicals, D-82377 Penzberg, Germany

Received March 31, 1994; Revised and Accepted July 14, 1994

ABSTRACT

Fluorescently labeled deoxynucleotides were used for internal labeling of DNA sequencing fragments generated in a two-color peak-height encoded protocol. Sequenase and a manganese-containing buffer were used to generate uniform peak heights. Tetramethyl rhodamine – dATP was used in a labeling step, followed by termination with ddATP and ddCTP in a 3:1 ratio. Fluorescein – dATP was used in a second reaction, followed by termination with ddGTP and ddTTP in a 3:1 ratio. The fragments were pooled and separated by capillary gel electrophoresis. The results were compared with peak-height encoded sequencing based on fluorescently labeled primers. The dye-labeled primers produced higher resolution separations for shorter fragments. However, dye-labeled primer fragments suffered from an earlier onset of biased reptation and produced shorter sequencing reads. Fragments from 50 to at least 500 bases in length could be sequenced with the internal labels.

INTRODUCTION

Current automated DNA sequencing technology makes use of fluorescent labels to detect and identify the DNA fragments. The labeled fragments are generated either by use of labeled primers or by incorporation of labeled dideoxynucleotide triphosphates (ddNTPs) (1–3). In 1992 Voss *et al.* reported the use of both fluorescein-12–dUTP (F-12–dUTP) and fluorescein-15–dATP (F-15–dATP) as internal labels for DNA sequencing with T7 DNA polymerase and demonstrated this labeling strategy for primer walking applications (4–5). More recently, Hou and Smith demonstrated hexamer priming with internal fluorescence labeling (6).

The use of fluorescently tagged deoxynucleotides presents three advantages over the current labeling techniques. First, the incorporation of the fluorescent dNTP in a separate labeling step is compatible with a wide range of primers, hence time consuming and costly preparation of fluorescently labeled primers for primer walking is avoided. Second, the incorporation of these dNTP labels is quite uniform. Extensions and terminations performed with T7 DNA polymerase and manganese demonstrate uniform peak intensities, which increase the accuracy of the sequence. Uniform terminations are not obtained, even under optimized experimental conditions, with the labeled ddNTPs (3). Third, the available dye-labeled deoxynucleotides are much less expensive than fluorescently labeled dideoxynucleotides or fluorescently labeled primers.

In 1992 we reported the two-label peak-height encoded DNA sequencing technique (7–8). This method relies on the uniform incorporation of ddNTPs achieved by T7 DNA polymerase in the presence of manganese (9). The sequencing fragments are separated on a single lane of a slab gel or in a single gel-filled capillary tube. Samples labeled with fluorescent primers demonstrate the uniform peak heights required for this method.

For the two-label peak-height encoded technique to be useful with capillary gel electrophoresis, two spectrally different labels are required. Until recently, only fluorescein-labeled dNTPs have been available. Boehringer Mannheim have made available to us a tetramethylrhodamine labeled dATP (TR–dATP). Simple modification of the previously reported detection system allows detection of both the fluorescein and the tetramethylrhodamine labels. This paper describes the use of internal labeling with the two-label peak-height encoded sequencing technique and compares the separation of these samples with that of primer-labeled samples.

MATERIALS AND METHODS

Instrumental design, electrophoresis

The system used for DNA sequencing by capillary electrophoresis is similar to those previously reported (7,10). The polyimide-coated fused-silica capillary tubing is 50 μm inner diameter, 145 μm outer diameter, and is typically 35 cm long. The non-crosslinked polyacrylamide is prepared from 5 ml aliquots of carefully degassed acrylamide solution (6%T), 1×TBE, and 7 M urea. Polymerization is initiated by addition of 2 μl of

*To whom correspondence should be addressed

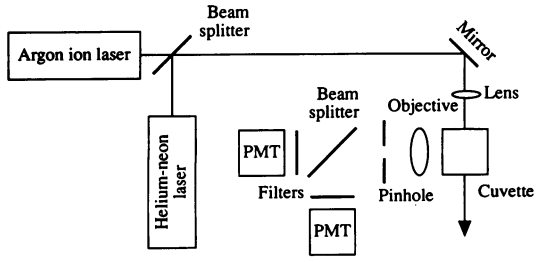


Figure 1. Laser-induced fluorescence detector. Beams from an argon ion laser and a helium–neon laser are combined with a dichroic beam splitter, reflected from a mirror, and focused with a lens into a sheath flow cuvette. Fluorescence is collected at right angles with a microscope objective and imaged onto a pinhole. The transmitted light is divided into two spectral channels with a dichroic beam splitter. Light in each channel is spectrally filtered with a bandpass filter and detected with a photomultiplier tube.

TEMED and 20 μl of 10% ammonium persulfate. The capillaries are treated with a solution of γ-methacryloxypropyltrimethoxysilane to bind the gel to the capillary walls. The capillaries are filled with the acrylamide monomer solution by application of a vacuum. Polymerization occurs within the capillary and appears complete within 30 min, however, capillaries are typically stored overnight before use.

The injection end of the capillary is held inside a Plexiglas box equipped with a safety interlock. The other end of the capillary is inserted into the flow chamber of a locally constructed sheath flow cuvette (7,10–11). The cuvette has a 150 μm square flow chamber and 1 mm thick quartz windows.

Detector

The fluorescence detector is shown in Figure 1. A 10 mW argon ion laser beam (λ = 488 nm) is aligned with a dichroic beam splitter to be coincident with a 5 mW helium–neon laser beam (λ = 543.5 nm). The combined laser beams are focused with a 5× microscope objective onto the sheath flow cuvette about 100 μm below the tip of the capillary. Fluorescence is collected at a right angle to the laser beams with a 0.6 NA, 32× microscope objective (Leitz/Wild model 2569–1130). The fluorescence is imaged onto a 0.75 mm diameter pinhole. A dichroic beam splitter divides the fluorescence into two spectral channels. The transmitted fluorescence passes through a spectral bandpass filter centered at 580 nm with a 30 nm bandwidth; this spectral channel primarily detects tetramethylrhodamine. The reflected fluorescence passes through a spectral bandpass filter centered at 515 nm with a 25 nm bandwidth; this spectral channel detects fluorescein. In both channels the fluorescence is detected with an R1477 photomultiplier tube (PMT). The output from each PMT is conditioned with a low pass electronic filter with a time constant of 0.5 s, and digitized by a National Instruments A/D board in a Macintosh IIsi computer. Data is collected at a sampling rate of 2 Hz. Before presentation, the data are smoothed with a Gaussian filter and processed to compensate for a mobility shift. The sheath flow is 1×TBE provided by a simple siphon based on a 4 cm height difference between the sheath buffer reservoir and the waste collection vial. The sheath flow rate is approximately 0.08 ml/h.

Sample preparation

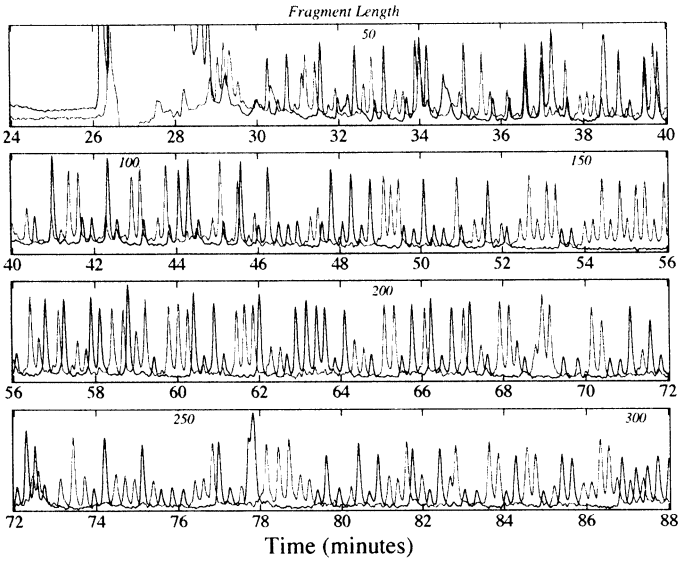
The sequencing reactions were carried out in buffers from the Pharmacia AutoRead sequencing kit. The G and T terminations were prepared by mixing 1 μg of M13mp18 single stranded DNA with 2 μl of −40 primer (5 μM), 2 μl of annealing buffer (1 M Tris–HCl, pH 7.4, 100 mM $MgCl_2$), and H_2O to a volume of 15 μl. Annealing was accomplished by heating the sample to 65°C for 2 min, followed by slow cooling to room temperature. The labeling reaction was carried out by addition of 1 μl of extension mix (40 mM $MnCl_2$, 302 mM Na-citrate, 324 mM DTT), 2 μl labeling mix (10 μM F-15–dATP, 1 μM each dCTP, dGTP, and dTTP), and 6 units of Sequenase Version 2.0 (United States Biochemical, Cleveland, OH). The reaction was incubated at 37°C for 10 min. The labeling reaction was followed by addition of 3.5 μl of DMSO and 10 μl of the combined G and T termination mixes (mixed in a 3:1 ratio G:T; each mix is 1 μM of each dATP, dCTP, dGTP, dTTP and 5 μM of the respective ddNTP, 50 mM NaCl, 40 mM Tris–HCl, pH 7.4). The termination reaction was carried out at 37°C for 10 min. The reactions were stopped by addition of 13 μl of stop solution (1 M NaOAc, pH 8.0, 20 mM EDTA) and the DNA was precipitated with ethanol. Identical conditions were used to generate the tetramethylrhodamine-labeled sample. The labeling mix contained 10 μM TR–dATP and the A and C termination mixes were added in a 3:1 ratio. The samples were each resuspended in 3 μl of a 49:1 mixture of formamide–0.5 M EDTA. The F-15–dATP-labeled sample was mixed with the TR–dATP-labeled sample and heated to 95°C for 2 min. The sample was injected onto the capillary by applying a 200 V/cm electric field for 60 s. The sample was separated in an electric field of 200 V/cm.

The comparison study was performed in an electric field of 300 V/cm. This higher field strength accentuates the effects of biased reptation. For the comparison, ddC-terminated M13mp18 samples were prepared in the manner described above. The internally labeled sample was labeled with F-15–dATP. The primer-labeled sample was labeled with a locally synthesized FAM-labeled −40 M13 primer, and the labeling step was eliminated from this synthesis.

RESULTS AND DISCUSSION

Two lasers are used in the system described. Tetramethylrhodamine is excited by the green helium–neon laser operating at 543.5 nm and, to a lesser degree, by the argon ion laser operating at 488 nm. Fluorescein is excited only by the argon ion laser. The argon ion laser is operated at 10 mW to minimize the background due to Raman scatter from water in the transmitted channel centered at 580 nm, while simultaneously enhancing the signal from tetramethylrhodamine. Fluorescein is photobleached at this laser power.

The sequencing electropherogram is shown in Figure 2. The red trace (transmitted channel) shows the A and C terminations; the large amplitude peaks are A, the small are C. The blue trace (reflected channel) shows the G and T terminations; large peaks are G, small are T. The data was processed as described below. There is a small amount of cross-talk in the transmitted channel; some fluorescence from the F-15–dATP is transmitted by the 580 nm bandpass filter. A mobility shift between the two channels was also observed; the tetramethylrhodamine-labeled fragments, detected in the transmitted channel, migrated faster than the fluorescein-labeled fragments, detected in the reflected channel. This shift is presumably due to differences in the mobilities of the labels. The mobility shift was corrected by normalizing the two channels, subtracting the small contribution from optical



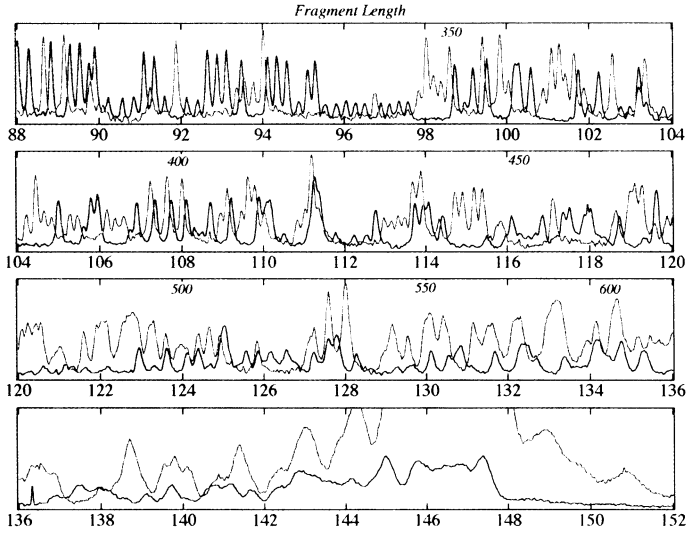


Figure 2. Capillary gel electropherogram of an M13mp18 template, sequenced with the two-color peak-height encoded technique. Fluorescein−dATP and tetramethyl rhodamine−dATP were used to label the fragments. Retention time in minutes is shown below and fragment length (including primer) is shown at the top of each panel. Experimental conditions are shown in the text.

cross-talk in the transmitted channel, and shifting the transmitted channel data back by 10 s.

The DNA sequence may be identified for fragments 50−500 bases long. In this region, there were a total of 12 errors, corresponding to a sequencing accuracy of 97%. A major compression at base 420 resulted in the loss of 6 bases of sequence. Of the remaining errors, five were due to missing Cs and one was due to a missing G. Two of the miscalled Cs were due to overlap with neighboring As. Two Cs had coalesced and were miscalled as an A. The miscalled G arose because two peaks had coalesced. It is clear that the denaturation capability of the gel, when operated at room temperature, leads to most of the sequencing errors.

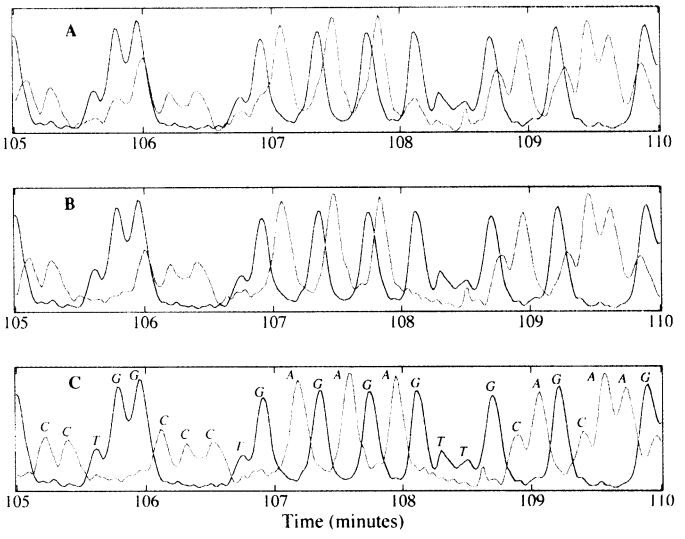


Figure 3. Demonstration of data treatment. (**A**) The filtered data. (**B**) The data after correction for spectral overlap. (**C**) The data after correction for mobility shift. The called sequence is shown above each peak in panel C.

If the data is not processed to compensate for the mobility shift, peaks corresponding to Cs (small peaks in the solid trace) are frequently missed when they immediately follow a G (large peaks in the slower dashed trace). This effect is illustrated in Figure 3. The top panel shows the smoothed data. The middle panel shows the data after correction for spectral cross-talk; the red trace was replaced by the original value minus 0.25×the blue trace. The bottom panel shows the data after correction for mobility shift. The red channel was offset by 7 s. The bottom panel also includes the observed sequence.

To investigate further the extent of incorporation of the labeled nucleotide, a reaction was stopped after the labeling step. Depending on the experimental conditions, the primer is extended 6−8 nucleotides to a quartet of As. The main product (80−90%) of this labeling reaction incorporates one labeled dATP. Minor products, incorporating two or three labeled dATPs are observed. Based on these results, an investigation was undertaken to determine the effect of the internal labeling on the efficiency of the electrophoretic separation. The incorporation of more than one labeled nucleotide in fragments of a given length could cause the fragments to have a larger distribution of mobilities and introduce peak broadening.

Two samples, one primer-labeled, the other internally labeled, were run on separate, identical capillary gels in an electric field of 300 V/cm. The gels were prepared from the same batch of acrylamide and were stored for the same length of time before the runs. During the experiment the lengths of the capillaries were the same within a few millimeters, and the pre-run times were the same to within 3 min. The two runs were carried out in series, over a 6 h period. This procedure should present negligible differences in the gel composition inside the capillaries, and so the two runs took place under essentially identical conditions, facilitating comparison of the results.

A plot of the retention times versus the fragment length is shown in Figure 4a. The internally labeled fragments have longer retention times, hence lower mobilities than the primer-labeled

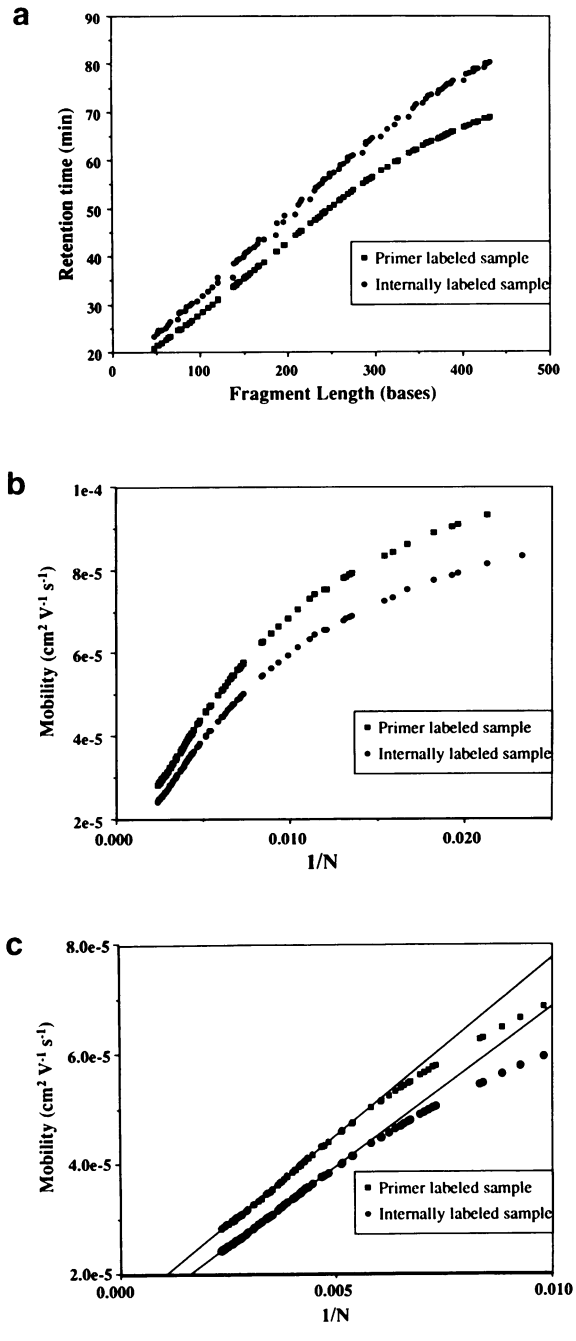


Figure 4. Evaluation of biased reptation. (**a**) The retention time for both internally labeled and primer-labeled DNA sequencing fragments. (**b**) The mobility of the sequencing fragments versus inverse fragment length ($1/N$). (**c**) A linear fit to the mobility of longer fragments.

fragments of the same length. According to the biased reptation model, mobility is inversely proportional to fragment length

$$\mu = \chi\left[\frac{1}{N} + \frac{1}{N^*}\right] \quad (1)$$

where χ is a proportionality constant, N is the fragment length in bases, and N^* is the fragment length for the onset of biased reptation. Figure 4b presents plots of the inverse fragment length versus mobility. A straight line was fitted to the data for fragments

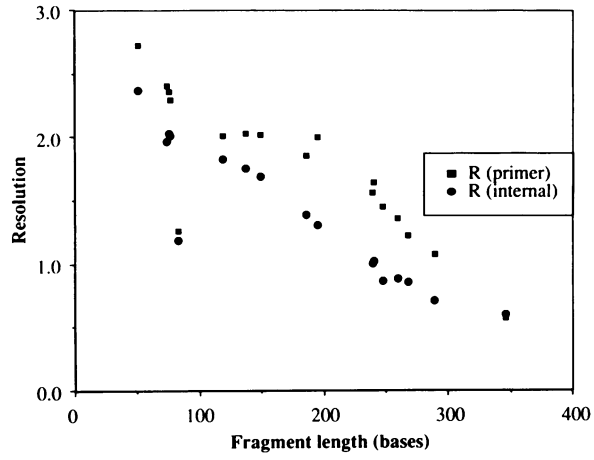


Figure 5. Resolution as a function of fragment length for both internally labeled and primer-labeled DNA sequencing fragments.

longer than 200 bases (Figure 4c); the onset of biased reptation for primer-labeled fragments was observed at 490 bases, while the internally labeled sample demonstrated an onset of biased reptation at 545 bases. The delayed onset of biased reptation for the internally labeled samples is likely responsible for the extremely long runs at low electric field reported by Ansorge and co-workers (5).

Chromatographers use resolution to describe the relative separation of adjacent peaks

$$R = \frac{\Delta T}{W_{\text{baseline}}} \quad (2)$$

where R is resolution, ΔT is peak spacing, and $\overline{W_{\text{baseline}}}$ is the peak-width measured at the baseline. The resolution of adjacent peaks of all multiplets in the electropherograms was measured. Figure 5 presents resolution as a function of fragment length. Several points can be made about this resolution data. First, resolution decreases steadily with fragment length. Second, a compression is observed for fragments about 85 bases long. This compression results in a anomalously low resolution for both labeling schemes. Third, there appears to be a constant 0.4 difference in resolution between the two labeling schemes for fragments up to 300 bases in length. This consistently poorer resolution observed for the internally labeled fragments presumably is caused by a distribution of numbers of labels incorporated in the sequencing fragments. Since the fluorescent label contributes to the mobility of the fragment, a distribution of labels will lead to a distribution of mobilities for fragments of identical length, leading to band broadening and degraded resolution. Fourth, the resolution for the two labeling techniques are similar for longer fragments. Here, the onset of biased reptation causes fragments with greater than 350 bases to crowd together to a larger extent for primer-labeled fragments compared with internally labeled fragments.

CONCLUSIONS

The incorporation of the fluorescently labeled deoxynucleotides into DNA sequencing samples for the two-label peak-height encoded technique is quite readily achieved. There is a mobility

difference caused by the use of two different labels. This may be compensated for by suitable processing of the raw data. After minimal data processing, sequence may be obtained from fragments at least 500 bases long.

However, there are several potential limitations to this technology. First, the technique requires use of T7 DNA polymerase. Until a thermally stable polymerase is discovered that produces uniform incorporation of dideoxynucleosides, cycle sequencing and PCR-based sequencing will remain impractical with the peak-height encoded sequencing technique. Second, we chose a sequencing primer so that the extension reaction terminated at a quartet of As. We verified that sequence is generated with a −21 primer and M13mp18 template. Voss also has reported internal labeling for primer walking experiments (4−5). However, we have not verified that internal labeling and peak-height encoding work well for a wide range of primers in primer walking experiments.

The fluorescein−tetramethylrhodamine labeling pair suffers from high background in the tetramethylrhodamine channel. The spectral filter for the tetramethylrhodamine transmits water Raman scatter caused by the argon ion laser used to excite the fluorescein. Raman scatter degrades the signal-to-noise ratio for the tetramethylrhodamine label; this label already gives less signal than the fluorescein. Other label pairs, for example, tetramethylrhodamine and Texas Red, or fluorescein and 5-carboxy-4′,5′-dichloro-2′,7′-dimethoxyfluorescein would be better suited to this method. These dye pairs could be excited with one laser, the green helium−neon or the blue argon ion, respectively, thus simplifying the instrument design. More importantly, the detection systems could be designed to minimize background and improve the signal-to-noise ratio for the labels, thus improving detection of small amounts of DNA.

ACKNOWLEDGEMENTS

This work was supported in part by the Department of Energy−Human Genome Initiative (USA) grant number DE-FGO2−91ER61123. Support by DOE does not constitute an endorsement of the views expressed in this article. This work was also supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Department of Chemistry of the University of Alberta, and Boehringer Mannheim. H.R.S. acknowledges a predoctoral fellowship from the Alberta Heritage Foundation for Medical Research. N.J.D. acknowledges a Steacie fellowship from NSERC.

REFERENCES

1. Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B.H., Hood, L.M. (1986), *Nature* **321**, 674−679
2. Prober, J.M., Trainor, G.L., Dam, R.J., Hobbs, F.W., Robertson, C.W., Zagursky, R.J., Cocuzza, A.J., Jensen, M.A., Bauermeister, K. R. (1987) *Science* **238**, 336−341.
3. Lee, L.G., Connell, C.R., Woo, S.L., Cheng, R.D., McArdle, B.F., Fuller, C.W., Halloran, N.D., Wilson, R.K. (1992) *Nucleic Acids Res.* **20**, 2471−2483.
4. Voss, H., Schwager, C., Wirkner, U., Zimmermann, J., Erfle, H., Hewitt, N.A., Rupp, T., Stegemann, J., Ansorge, W. (1992) *Methods Mol. Cell. Biol.* **3**, 30−34.
5. Voss, H., Wiemann, S., Wirkner, U., Schwager, C., Zimmermann, J., Stegemann, J., Erfle, H., Hewitt, N.A., Rupp, T., Ansorge, W. (1992) *Methods Mol. Cell. Biol.* **3**, 153−155.
6. Hou, W., Smith, L.M. (1993) *Nucleic Acids Res.* **21**, 3331−3332.
7. Chen, D.Y., Harke, H.R., Dovichi, N.J. (1992) *Nucleic Acids Res.* **20**, 4873−4880.
8. Lu, H., Arriaga, A., Chen, D.Y., Dovichi, N.J. (1994) *J. Chromatogr.* in press.
9. Tabor, S., Richardson, C.C. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 4076−4080.
10. Swerdlow, H.P., Zhang, J.Z., Chen, D.Y., Harke, H.R., Grey, R., Wu, S., Fuller, C., Dovichi, N.J. (1991) *Anal. Chem.* **63**, 2835−2841.
11. Cheng, Y.F., Dovichi, N.J. (1988) *Science* **242**, 562−564
12. Dovichi, N.J., Martin, J.C., Jett, J.H., Keller, R.A. (1983) *Science* **219**, 845−847.